A General Framework for the Design and Analysis of Sparse FIR Linear Equalizers

Abubakr O. Alabbasi.

Prof. Naofal Al-Dhahir.
Prof. Ridha Hamila.
Prof. Waheed Bajwa.

Exploiting Sparsity for Interference Management in Broadband Networks: Theory, Application, and Testbeds.
NPRP No. 06-070-2-024

GlobalSIP - Dec. 2015

Outlines

- *Background about Sparse Approximation(spA).*
- *Motivation and Related Work.*
- *Our Objectives(Contributions).*
- *Sparse FIR Linear Equalizers.*
- *Worst-Case Coherence Analysis*
- *Selected Results.*
- *Conclusions.*

Outlines

- ***Background about Sparse Approximation(spA).***
- *Motivation and Related Work.*
- *Our Objectives*
- *System Model*
- *Measurement Matrix Properties.*
- *Selected Results.*

What is SpA ?

- SpA is the problem of approximating a signal (vector) with the best linear m combination of elements from a redundant dictionary. Greedy is Good [Joel A. Tropp, 2006]

$$\begin{pmatrix} 0.9593 & 0.2575 & 0.2435 & 0.2511 & 0.8308 \\ 0.5472 & 0.8407 & 0.9293 & 0.6160 & 0.5853 \\ 0.1386 & 0.2543 & 0.3500 & 0.4733 & 0.5497 \\ 0.1493 & 0.8143 & 0.1966 & 0.3517 & 0.9172 \end{pmatrix} \begin{pmatrix} 0 \\ 0.7537 \\ 0 \\ 0 \\ 0 \end{pmatrix} \approx \begin{pmatrix} 0.2254 \\ 0.7531 \\ 0.2366 \\ 0.6390 \end{pmatrix}$$

Sparse Approximation Review

- Consider the linear system of Equations

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{z}$$

$$\boldsymbol{y} \in \mathbb{C}^M \qquad \boldsymbol{x} \in \mathbb{C}^N \qquad \boldsymbol{z} \in \mathbb{C}^M$$

Where:

A : represents the Sparsifying Dictionary.

y : represents the Data (MD).

x : represents the Original Signal.

z : Additive Noise.

- Two main approaches have been proposed in the literature to solve the spA problems; specifically,
  - ❑ $\boldsymbol{l_1 -}$ **minimization.**
  - ❑ **Greedy algorithms.**

Sparse Approximation Review

1. **Convex optimization**: l_1-minimization.

$$\min_{\tilde{\boldsymbol{x}} \in \mathbb{C}^N} \| \tilde{\boldsymbol{x}} \|_0 \quad \text{s.t.} \quad \| \boldsymbol{y} - \boldsymbol{A}\tilde{\boldsymbol{x}} \|_2^2 \leq \epsilon$$

2. **Greedy algorithms**: One of the most widely greedy algorithms is the **O**rthogonal **M**atching **P**ursuit (OMP)

$$\tilde{x} = \text{OMP}\ (y, A, \text{stopping criterion})$$

↓

Orthogonal **M**atching **P**ursuit Algorithm

OMP: Selects the most corr. columns of (A) with the MD (y).

Other Algorithms:

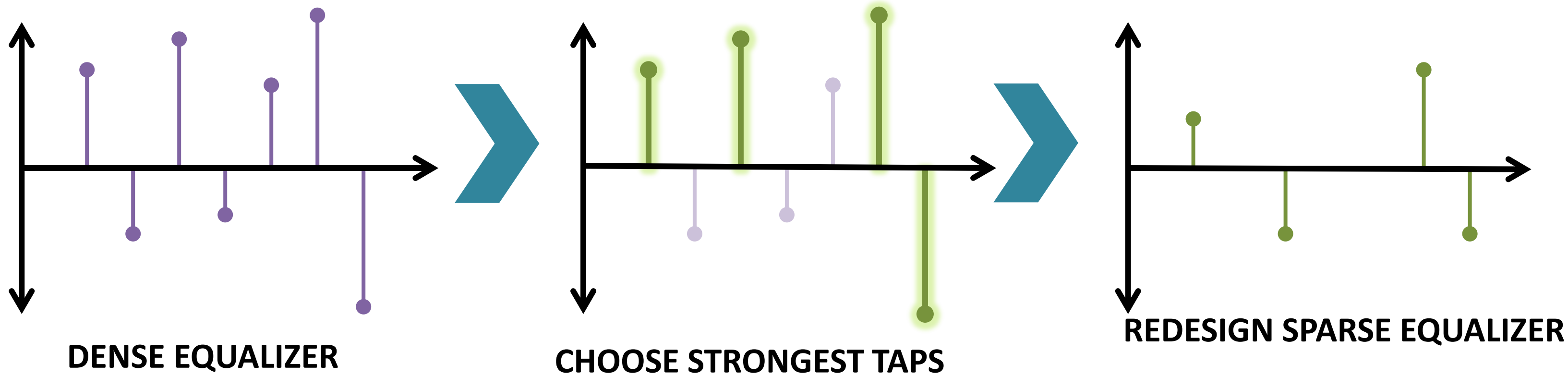
- **LASSO**
- **OLS**
- **StOMP**
- **ROMP**
- **CoSaMP**
- **IHT**

Outlines

- Background about Compressive Sensing(CS).
- **Motivation and Related Work.**
- Our Objectives
- System Model
- Measurement Matrix Properties.
- Selected Results.

Motivation and Related Work

- Highly dispersive comm. Channels are ch'c by IR's that span tens-hundreds of symbol periods.
- Very long equalizers have to be employed at the Rx to mitigate the resulting sever ISI. **[Bingham, 1990].**
- Complexity of finite-impulse-response (FIR) equalizers is proportional to the square of the number of nonzero taps in the filter**.**



FIR** filters with non-consecutive non-zero taps typically referred to as **sparse filters

'Strongest tap selection' method **[Melvasalo, 2007] [Melvasalo, 2010] [Gomaa, 2012]**

1. Design large length dense filter and choose a subset of strongest taps.
2. Design sparse filter on the selected locations.

Outlines

- Background about Compressive Sensing(CS).
- Motivation and Related Work.
- ***Our Objectives***
- System Model
- Measurement Matrix Properties.
- Selected Results.

Our Contributions,

- *Designing a general framework that transforms the problem of design of sparse linear equalizers (**LEs**) into the problem of sparsest approximation of a vector in different dictionaries.*

Outlines

- Background about Compressive Sensing(CS).
- Motivation and Related Work.
- Our Objectives
- ***Sparse FIR LE problem formulation***
- Measurement Matrix Properties.
- Selected Results.

Sparse Equalization

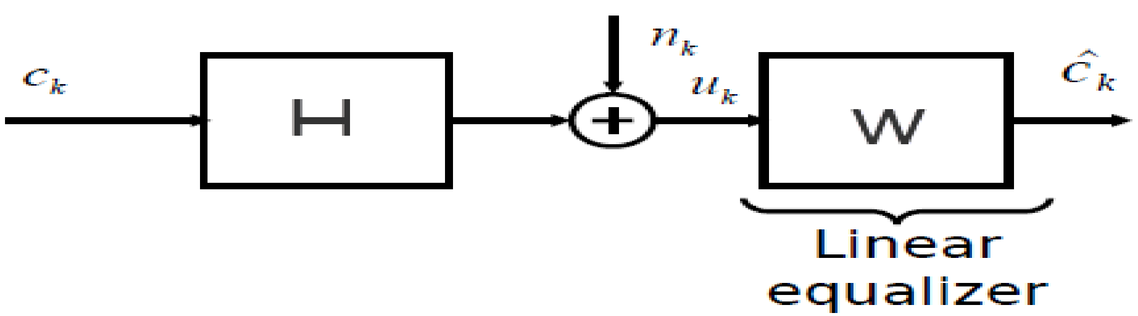
- We consider a general case of a digital communication system with n_i inputs and n_o outputs.
- The samples at the j^{th} output $(1 \leq \mathrm{j}^{\mathrm{th}} \leq \mathrm{n_o})$

$$\boldsymbol{y}_k^{(j)} = \sum_{i=1}^{n_i} \sum_{m=0}^{\nu^{(i,j)}} \boldsymbol{h}_m^{(i,j)} x_{k-m}^{(i)} + \boldsymbol{n}_k^{(j)}$$

- Grouping over a block of N_f symbol periods,

$$\boldsymbol{y}_{k:k-N_f+1} = \boldsymbol{H}\boldsymbol{x}_{k:k-N_f-\nu+1} + \boldsymbol{n}_{k:k-N_f+1}$$

- **For SISO-LEs (Error Signal):**



$$e_k = x_{k-\Delta} - \hat{x}_k = x_{k-\Delta} - \boldsymbol{w}^H y_{k:k-N_f+1} \,,$$

Sparse FIR SISO-LEs

$$\text{MSE} = \xi = E\left[\left|e_k^2\right|\right]$$
$$= E\left[e_k^H e_k\right]$$
$$= E\left[\left(x_{k-\Delta} - \boldsymbol{w}^H y_{k:k-N_f+1}\right)^H \left(x_{k-\Delta} - \boldsymbol{w}^H y_{k:k-N_f+1}\right)\right]$$
$$= E\left[x_{k-\Delta}^2\right] - \boldsymbol{w}^H \underbrace{E\left[y_{k:k-N_f+1} x_{k-\Delta}\right]}_{\boldsymbol{R}_{yx}} - E\left[y_{k:k-N_f+1} x_{k-\Delta}\right]^H \boldsymbol{w} + \boldsymbol{w}^H \underbrace{\left[y_{k:k-N_f+1} y_{k:k-N_f+1}^H\right]}_{\boldsymbol{R}_{yy}} \boldsymbol{w}$$

$$\xi(\boldsymbol{w}) = \underbrace{\varepsilon_x - r_\Delta^H \boldsymbol{R}_{yy}^{-1} r_\Delta}_{\xi_m} + \underbrace{(\boldsymbol{w} - \boldsymbol{R}_{yy}^{-1} r_\Delta)^H \boldsymbol{R}_{yy} (\boldsymbol{w} - \boldsymbol{R}_{yy}^{-1} r_\Delta)}_{\xi_e(\boldsymbol{w})}$$

$$\widehat{w}_s \triangleq \underset{\boldsymbol{w} \in \mathbb{C}^{N_f}}{\text{argmin}} \; \|w\|_0 \quad \text{subject to} \quad \xi_e(\boldsymbol{w}) \leq \delta_{eq} \quad \text{(1)}$$

- W^H : Filter taps.
- ϵ_x: Signal Energy.
- $r_\Delta = R_{yx}^{-1} * 1_\Delta$, where 1_Δ is all zero vector except the (Δ+1) entry, where this entry is 1.
- W^H, R_{yx}, R_{yy} : Equalizer Coeff., Cross-Corr. Matrix, Output Auto-corr. Matrix.

Proposed sparse approximation framework

- **We provide a general *framework for designing sparse FIR LEs and equalizers* that can be considered as the problem of sparse approximation using different dictionaries. Mathematically, this framework poses the design problem as follows:**

$$\widehat{\boldsymbol{w}}_s \triangleq \underset{\boldsymbol{w}\in\mathbb{C}^{N_f}}{\operatorname{argmin}} \|\boldsymbol{w}\|_0 \quad \text{subject to} \quad \|\boldsymbol{A}(\boldsymbol{\Phi}\boldsymbol{w}-\boldsymbol{b})\|_2^2 \leq \delta_{eq}$$

where $\boldsymbol{\Phi}$ is the dictionary that will be used to sparsely approximate $\boldsymbol{w}$, while $\boldsymbol{A}$ is a known matrix and $\boldsymbol{b}$ is a known data vector, both of which change depending upon the sparsifying dictionary $\boldsymbol{\Phi}$.

Example:

Sparse FIR SISO-LEs

$$\min_{\boldsymbol{w}\in\mathbb{C}^{N_f}} \|\boldsymbol{w}\|_0 \quad \text{s.t} \quad \left\|\left(\boldsymbol{L}^H\boldsymbol{w} - \boldsymbol{L}^{-1}\boldsymbol{r}_\Delta\right)\right\|_2^2 \leq \delta_{eq}\,,$$

$$\min_{\boldsymbol{w}\in\mathbb{C}^{N_f}} \|\boldsymbol{w}\|_0 \quad \text{s.t} \quad \left\|\left(\boldsymbol{D}^{\frac{1}{2}}\boldsymbol{U}^H\boldsymbol{w} - \boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{U}^H\boldsymbol{r}_\Delta\right)\right\|_2^2 \leq \delta_{eq}, \text{ and}$$

$$\min_{\boldsymbol{w}\in\mathbb{C}^{N_f}} \|\boldsymbol{w}\|_0 \quad \text{s.t} \quad \left\|\boldsymbol{L}^{-1}\left(\boldsymbol{R}_{yy}\boldsymbol{w} - \boldsymbol{r}_\Delta\right)\right\|_2^2 \leq \delta_{eq}\,.$$

Table I

EXAMPLES OF DIFFERENT SPARSIFYING DICTIONARIES.

Cholesky Factorization			Eigen Decomposition		
$\boldsymbol{R}_{yy} = \boldsymbol{L}\boldsymbol{L}^H$ or $\boldsymbol{R}_{yy} = \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^H$			$\boldsymbol{R}_{yy} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^H$		
$\boldsymbol{A}$	$\boldsymbol{\Phi}$	$\boldsymbol{b}$	$\boldsymbol{A}$	$\boldsymbol{\Phi}$	$\boldsymbol{b}$
$\boldsymbol{I}$	$\boldsymbol{L}^H$	$\boldsymbol{L}^{-1}\boldsymbol{r}_\Delta$	$\boldsymbol{I}$	$\boldsymbol{D}^{\frac{1}{2}}\boldsymbol{U}^H$	$\boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{U}^H\boldsymbol{r}_\Delta$
$\boldsymbol{L}^{-1}$	$\boldsymbol{R}_{yy}$	$\boldsymbol{r}_\Delta$	$\boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{U}^H$	$\boldsymbol{R}_{yy}$	$\boldsymbol{r}_\Delta$
$\boldsymbol{I}$	$\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{P}^H$	$\boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{P}^{-1}\boldsymbol{r}_\Delta$	$\boldsymbol{D}^{\frac{1}{2}}$	$\boldsymbol{U}^H$	$\boldsymbol{D}^{-1}\boldsymbol{U}^H\boldsymbol{r}_\Delta$

To this end ?!

- ✓ **We have shown that the problem of designing sparse FIR equalizers can be cast into one of sparse approximation of a vector by a fixed dictionary. The general form of this problem is given by slide 23.**
- ✓ ***Which Sparsifying dictionary results in the sparsest design??***

Our next challenge is to determine the best sparsifying dictionary for use in our framework.

- **We know from the sparse approximation literature that the sparsity of the OMP solution tends to be inversely proportional to the worst-case coherence $\mu(\Phi)$.**

Worst-Case Coherence $\mu(\Phi)$ Analysis

- Once again, the sparsity of the OMP solution tends to be inversely proportional to the worst-case coherence

$$\mu(\boldsymbol{\Phi}) \triangleq \max_{i \neq j} \frac{\left|\left\langle \phi_i, \phi_j \right\rangle\right|}{\|\phi_i\|_2 \left\|\phi_j\right\|_2}$$

- Orthogonal Matching Pursuit (OMP) and like methods will estimate the optimum representation if $\mu(\emptyset) < 1$.[Tropp, 2006]

- Our Approach in investigating $\mu(\Phi)$?

$$\longrightarrow R_{yy} = HH^H + \frac{1}{SNR} I.$$

Worst-Case Coherence of R_{yy}, $\mu(R_{yy})$:

$$R_{yy} = HH^H + \frac{1}{SNR} I.$$

$$= \varepsilon_x \left[\begin{array}{c|cccccccc}
\overbrace{\sum_{i=0}^{v} |h_i|^2 + \frac{1}{\mathrm{SNR}}}^{a_0} & \overbrace{\sum_{i=1}^{v} h_i h'_{i-1}}^{a_{-1}} & \overbrace{\sum_{i=2}^{v} h_i h'_{i-2}}^{a_{-2}} & \cdots & \cdots & \overbrace{\sum_{i=v}^{v} h_i h'_{i-v}}^{a_{-v}} & 0 & \cdots & 0 \\
\hline
a_1 & a_0 & a_{-1} & a_{-2} & \cdots & \vdots & a_{-v} & 0\ldots & \cdots \\
a_2 & a_1 & a_0 & \ddots & \ddots & \vdots & & & \\
\vdots & \ddots & \ddots & a_0 & \ddots & \vdots & & & \\
\vdots & \vdots & \ddots & \ddots & \ddots & & & & \\
a_v & a'_{v-1} & \cdots & \cdots & \cdots & a_0 & & & \\
0 & \ddots & & \ddots & & & \ddots & & \\
\vdots & \cdots & \cdots & \cdots & \ddots & & & \ddots & \\
0 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & a_0
\end{array} \right]$$

$$R_{yy} = \mathrm{Toeplitz}\,\overbrace{\left(\begin{bmatrix} r_0 & r_1 & \ldots & r_v & 0 & \ldots & 0 \end{bmatrix} \right)}^{\phi_1^H},$$

Worst-Case Coherence of R_{yy}, $\mu(R_{yy})$:

$$R_{yy} = HH^H + \frac{1}{SNR} I.$$

$$\boldsymbol{R}_{yy} = \text{Toeplitz}\overbrace{\left(\left[\begin{array}{ccccccc} r_0 & r_1 & \dots & r_v & 0 & \dots & 0 \end{array}\right]\right)}^{\phi_1^H},$$

- **To get the worst CIR taps which result in the worst-case coherence:**

$$\max \sum_{i=1}^{v} |h_i h_{i-1}^*| \quad \text{s.t.} \quad \sum_{i=0}^{v} |h_i|^2 = 1.$$

$$\max \left| \boldsymbol{h}^H \boldsymbol{R} \boldsymbol{h} \right| \quad \text{s.t.} \quad \boldsymbol{h}^H \boldsymbol{h} = 1,$$

- **The solution of the above optimization problem is:**

$$\lambda_s = 2\cos\left(\frac{\pi s}{v+2}\right), \quad h_j^{(s)} = \sqrt{\frac{2}{v+2}} \sin\left(\frac{j\pi s}{v+2}\right),$$

Worst-Case Coherence of R_{yy} Factors

- **It is important to note here that the other dictionaries, which result from decomposing R_{yy} and $R_{x/y}$, can be considered as square roots of them in the spectral-norm sense. For example:**

$$\left\| \boldsymbol{R}_{yy} \right\|_2 = \left\| \boldsymbol{L}_y \boldsymbol{L}_y^H \right\|_2 \leq \left\| \boldsymbol{L}_y^H \right\|_2^2$$

$$\left\| \boldsymbol{R}_{\delta} \right\|_2 = \left\| \boldsymbol{U}_{\delta} \boldsymbol{D}_{\delta} \boldsymbol{U}_{\delta}^H \right\|_2^2 \leq \left\| \boldsymbol{D}_{\delta}^{1/2} \boldsymbol{U}_{\delta}^H \right\|_2^2$$

- **Through simulation, we show that the coherence of the factors of R_{yy} is less than that of $\mu(R_{yy})$.**

Can we do better?

Reduced-Complexity Design

- **The proposed designs for Sparse FIR LEs involve Cholesky factorization and/or eigen decomposition, whose computational costs could be large for channels with large delay spreads.**
- **For a Toeplitz matrix, the most efficient algorithms for Cholesky factorization are Levinson or Schur algorithms, which involve $O(M^2)$ computations, where *M* is the matrix dimension.**
- **In contrast, since a circulant matrix is asymptotically equivalent to a Toeplitz matrix, for reasonably large dimension, the eigen decomposition of a circulant matrix can be computed efficiently using the fast Fourier transform (FFT) and its inverse with only $O(Mlog(M))$ operations.**
- **By using this asymptotic equivalence between Toeplitz and circulant matrices, all computations needed for R_{yy} factorization can be done efficiently with the FFT and inverse FFT. In addition, direct matrix inversion can be avoided.**

Reduced-Complexity Design

- A circulant matrix, ***C***, has the discrete Fourier transform (DFT) basis vectors as its eigenvectors and the DFT of its first column as its eigenvalues.
- An $M \times M$ circulant matrix ***C*** can be decomposed as

$$C = \frac{1}{M}(F_M^H \Gamma_c F_M)$$

- The autocorrelation matrix R_{yy} is computed as:

$$\overline{\boldsymbol{R}}_{yy} = \underbrace{E\left[\widetilde{\boldsymbol{y}}_k \widetilde{\boldsymbol{y}}_k\right]}_{\overline{\boldsymbol{R}}_{\widetilde{y}\widetilde{y}}} + \underbrace{\frac{1}{SNR}}_{\sigma_n^2} \boldsymbol{I}_{N_f}$$

- To approximate R_{yy} as a circulant matrix, we assume that $\{y_k\}$ is cyclic. Hence, $E\{y_k y_k\}$ can be approximated as a time-averaged autocorrelation function as follows:

Reduced-Complexity Design: R_{yy}

$$
\begin{aligned}
\overline{\boldsymbol{R}}_{\widetilde{y}\widetilde{y}} &= \frac{1}{N_f}\sum_{k=0}^{N_f-1} \widetilde{\boldsymbol{y}}_k \widetilde{\boldsymbol{y}}_k^H = \frac{1}{N_f}\boldsymbol{C}_y \boldsymbol{C}_y^H \\
&= \frac{1}{N_f}\left(\frac{1}{N_f}\boldsymbol{F}_{N_f}^H \boldsymbol{\Lambda}_{\widetilde{Y}} \boldsymbol{F}_{N_f}\right)\left(\frac{1}{N_f}\boldsymbol{F}_{N_f}^H \boldsymbol{\Lambda}_{\widetilde{Y}}^H \boldsymbol{F}_{N_f}\right) \\
&= \frac{1}{N_f^2}\boldsymbol{F}_{N_f}^H \boldsymbol{\Lambda}_{\widetilde{Y}} \Lambda_{\widetilde{Y}}^H \boldsymbol{F}_{N_f} \\
&= \frac{1}{N_f^2}\left(\boldsymbol{F}_{N_f}^H \boldsymbol{\Lambda}_{|\widetilde{Y}|^2} \boldsymbol{F}_{N_f}\right).
\end{aligned}
$$

Note that $\boldsymbol{F}_{N_f}^H \boldsymbol{F}_{N_f} = \boldsymbol{F}_{N_f} \boldsymbol{F}_{N_f}^H = N_f \boldsymbol{I}_{N_f}$ and

$$
\boldsymbol{C}_y = \begin{bmatrix}
\widetilde{y}_{N_f-1} & \widetilde{y}_0 & \widetilde{y}_1 & \cdots & \widetilde{y}_{N_f-2} \\
\widetilde{y}_{N_f-2} & \widetilde{y}_{N_f-1} & \widetilde{y}_0 & & \\
\widetilde{y}_{N_f-3} & \widetilde{y}_{N_f-2} & \widetilde{y}_{N_f-1} & & \vdots \\
\vdots & & & \ddots & \\
\widetilde{y}_0 & & \cdots & & \widetilde{y}_{N_f-1}
\end{bmatrix}
$$

Reduced-Complexity Design: R_{yy}

$$\begin{aligned}
\widetilde{Y} &= H^H \odot P_\Delta \odot X \\
P_\Delta &= \begin{bmatrix} 1 & e^{-j2\pi\Delta/N_f} & \ldots & e^{-j2\pi(N_f-1)\Delta/N_f} \end{bmatrix}^T \\
\left|\widetilde{Y}\right|^2 &= N_f \left|H_i\right|^2 = N_f \begin{bmatrix} |H_0|^2 & \ldots & |H_{N_f-1}|^2 \end{bmatrix}^H
\end{aligned}$$

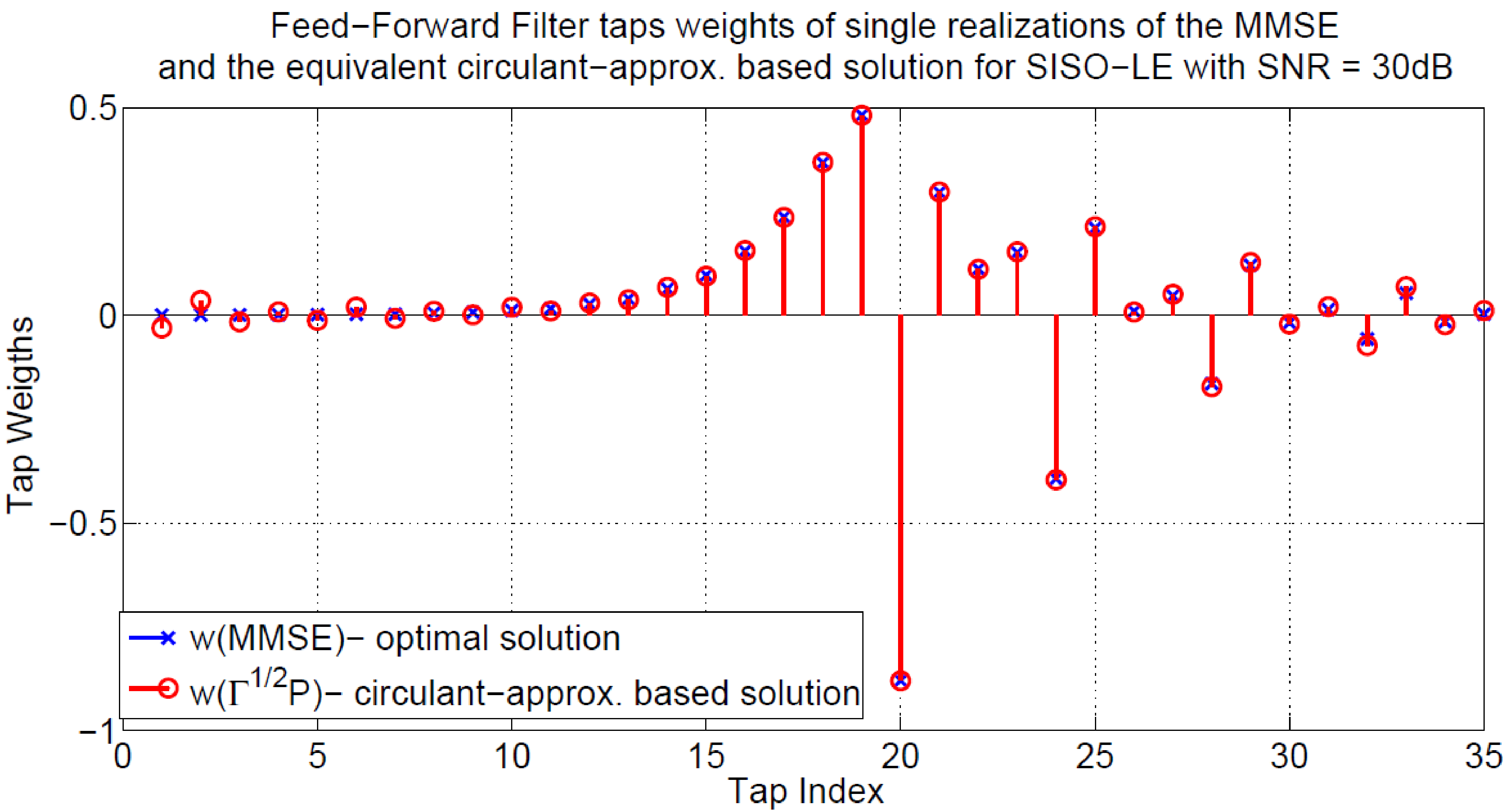
where $\odot$ denotes the element-wise multiplication. Then,

$$\begin{aligned}
\overline{\boldsymbol{R}}_{yy} &= \overline{\boldsymbol{R}}_{\tilde{y}\tilde{y}} + \sigma_n^2 \boldsymbol{I}_{N_f} \\
&= \frac{1}{N_f^2} \boldsymbol{F}_{N_f}^H \left(\boldsymbol{\Lambda}_{N_f|H|^2 + N_f \sigma_n^2 \mathbf{1}_{N_f}} \right) \boldsymbol{F}_{N_f} \\
&= \boldsymbol{F}_{N_f}^H \left(\frac{1}{N_f} \boldsymbol{\Lambda}_{\underbrace{|H|^2 + \sigma_n^2 \mathbf{1}_{N_f}}_{\varrho}} \right) \boldsymbol{F}_{N_f} \\
&= \boldsymbol{F}_{N_f}^H \left(\frac{1}{N_f} \boldsymbol{\Lambda}_{\varrho} \right) \boldsymbol{F}_{N_f} = \boldsymbol{Q}\boldsymbol{Q}^H .
\end{aligned}$$

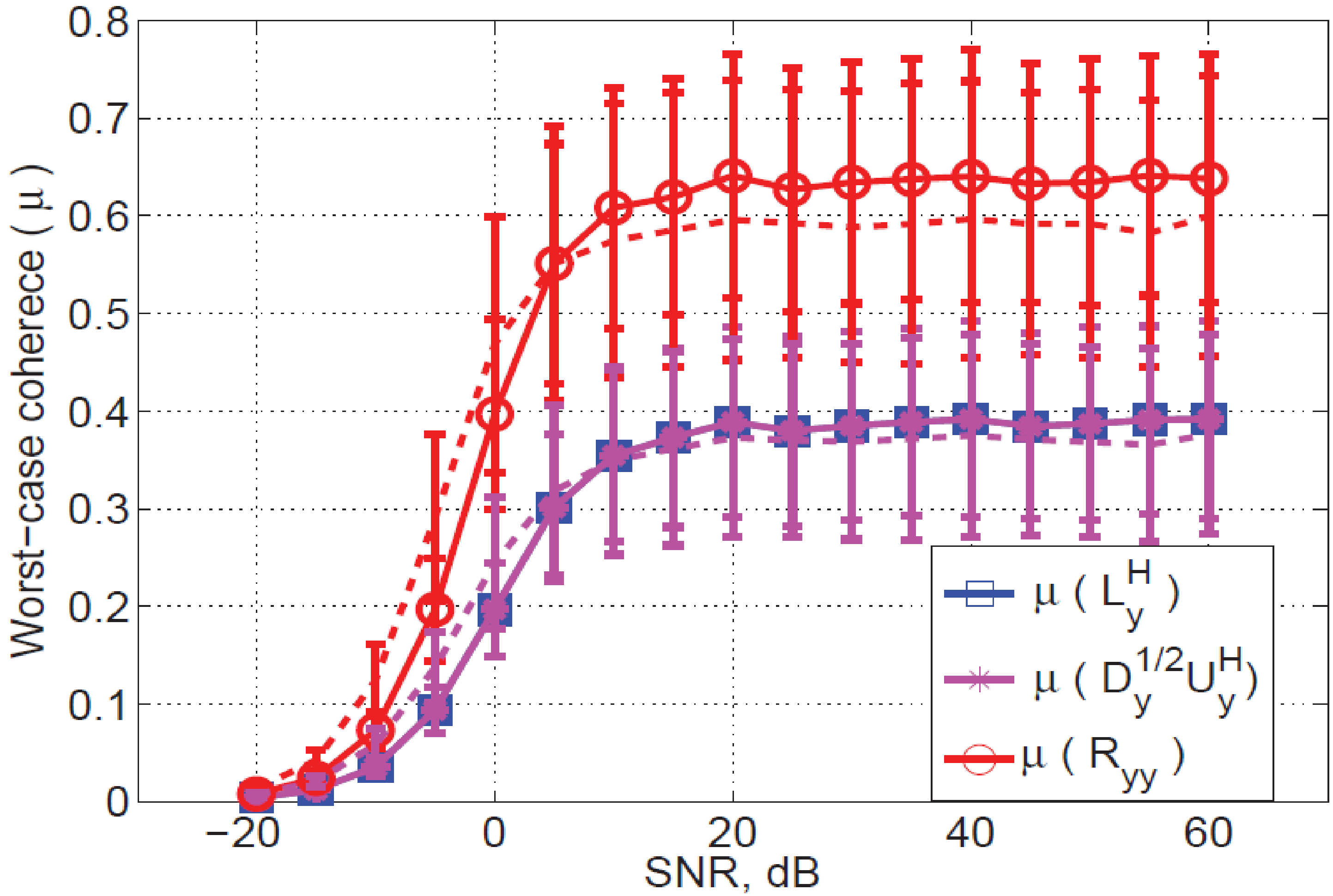
Outlines

- Background about Compressive Sensing(CS).
- Motivation and Related Work.
- Our Objectives
- Measurement Matrix Properties.
- ***Selected Results.***

Performance of circulant approximation based approach for UPDP channel with v = 5 and input SNR = 30dB.

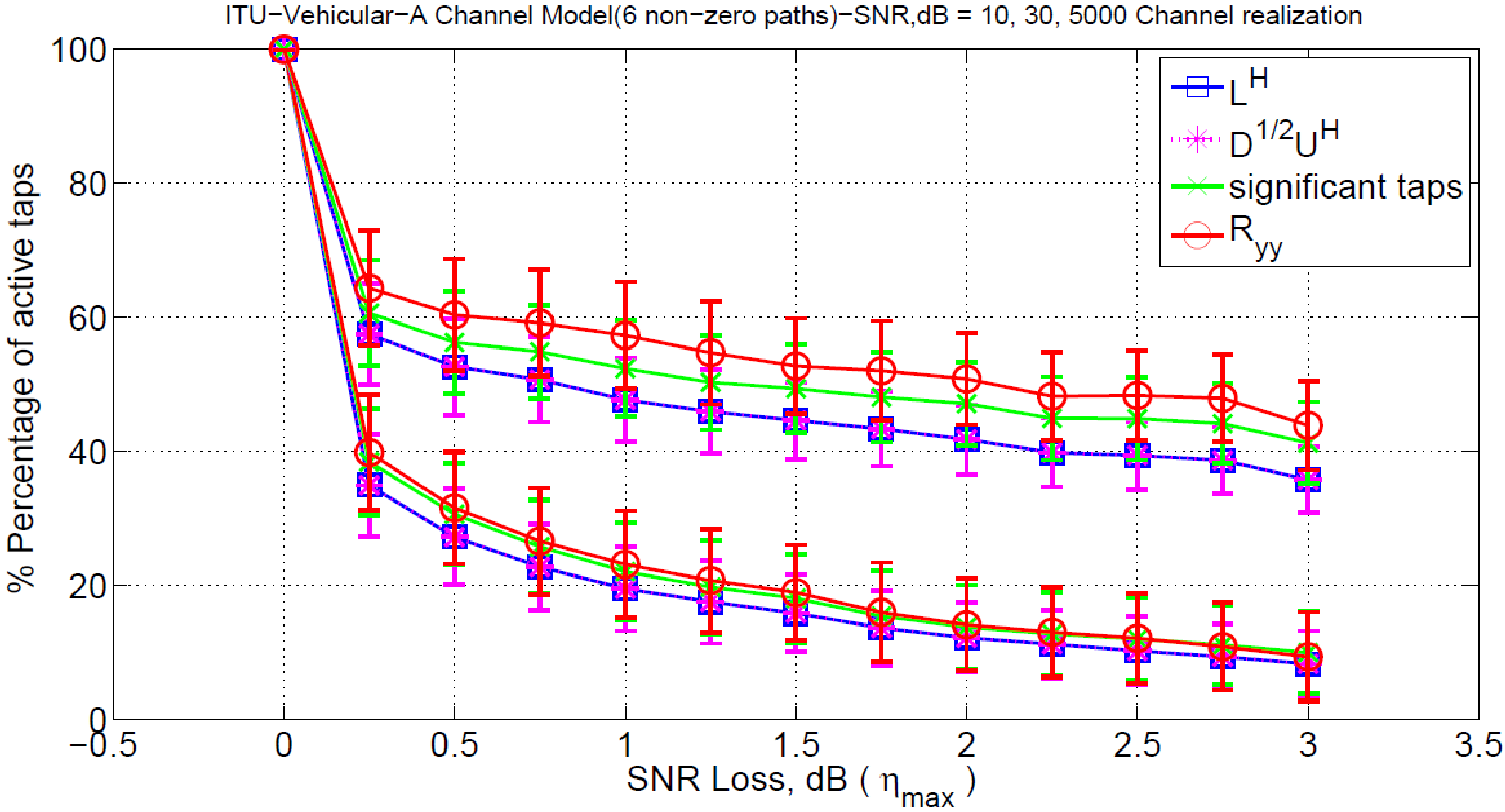


Worst-case coherence for R_{yy} and its factors L_y^H and $D_y^{\frac{1}{2}} U_y^H$



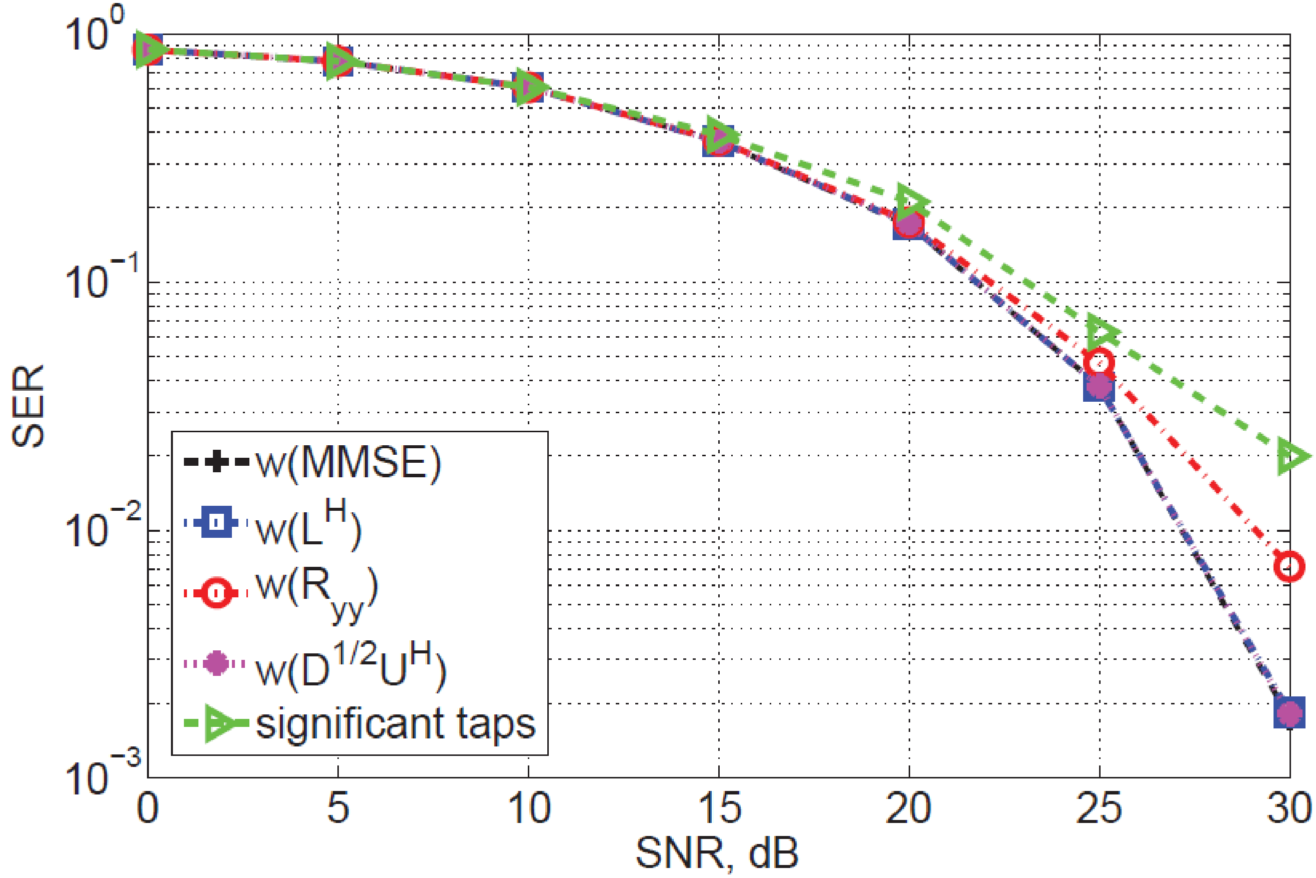
- ❑ Dashed lines represent the coherence of the corresponding circulant approximation for $D_y^{\frac{1}{2}} U_y^H$ (i.e., Q^H) and R_{yy} (i.e., $R_{yy} = QQ^H$).
- ❑ UPDP with $v = 8 \; and \; N_f = 80$.

Active taps Percentage versus Performance Loss: Vech-A-PDP (LEs)



- **Allowing for 0.25 dB performance loss results in a significant reduction in the number of the total filter taps (65% reduction) ☺☺**

SER versus SNR (LEs)



- **65% reduction (0.25dB loss) in FIR taps with almost same performance ☺☺**

Outlines

- ***Conclusions***

Conclusions

- Dramatic complexity reduction is achieved at a small performance loss compared to the conventional MMSE non-sparse FIR equalizer design.
- Sparse FIR equalizers enjoy more rapid taps weight adaptation to changing to channel condition which support higher mobile velocities.
- Based on the asymptotic equivalence of Toeplitz and circulant matrices, we provide reduced-complexity designs, where matrix factorizations can be carried out efficiently using the FFT and inverse FFT.
- The dictionary with the smallest coherence gives the sparsest filter design.

Thank you

a.o.alabbasi@ieee.org